

# Using Whole-Exome Imputation to Identify New Genetic Associations

September 15, 2014

JM Kocarnik

Adult body height is an easily-measured quantitative trait that serves as a model for studying the genetic architecture of complex polygenic traits. While genome-wide association studies (GWAS) have identified hundreds of height-associated variants, these loci only account for roughly 10 percent of the variation in height. While environmental exposures certainly play a role in adult body height, unmeasured rare and less-common variation is expected to account for much of the as-yet unexplained genetic component to height. Imputation of genetic variants using existing data is a cost-effective method for assessing some of these rarer genetic variants. In a recent report in *Human Molecular Genetics*, Drs. Margaret Du and Ulrike Peters and colleagues in the Public Health Sciences Division used whole-exome imputation to identify and replicate several independent loci associated with height in African Americans.

Genetic variants that are less frequent than those measured in GWAS are expected to have larger effects, but sequencing large numbers of people is expensive. Furthermore, the rarer the variant, the larger the study population needed to detect an association. This is problematic since sequencing becomes prohibitively expensive as the sample size increases. With limited resources, researchers must then balance the level of genomic coverage they are interested in investigating with the number of participants they can afford to include.

To partially address this limitation, a hybrid approach can be used that combines the deeper coverage of exome sequencing with the larger sample size of GWAS. In this study, the authors had access to GWAS data on nearly 14,000 African American participants from the NHLBI CARE and WHI-SHARE consortia. Rather than additionally sequence these samples, they utilized existing exome sequencing data from roughly 2,000 participants in the NHLBI Exome Sequencing Project. Using the shared variants as a guide, the fuller data from the exome-sequenced samples could then be imputed into the samples with GWAS data. Said lead author Dr. Du, "this increases the effective number of individuals with high quality exome sequence data and in turn increases statistical power."

While this method has been used successfully for other traits, "our study is the first to use exome imputation of sequence variants to investigate genetic variants related to height," said Du. After

imputation, the authors were able to investigate nearly 200,000 genetic variants in the exome for an association with height. Promising findings from this discovery set were then carried forward to an independent replication set of 2,000 African Americans with whole-exome sequence data.

Overall the authors identified and replicated eight variants in three independent loci, two of which demonstrated a novel association with height in African Americans (rs17410035 in *C5orf22* and rs114089985 in *SPRYD7*). The latter of these appears to be a population-specific allele that is infrequent in African Americans (3% minor allele frequency), very rare in European populations (0.03% minor allele frequency), and monomorphic in Asian populations. Despite having a large effect size (1.46 cm increase in height per A allele), this variant had not been found in previous GWAS in African Americans. Said Du, "our work provides proof of principle that whole-exome imputation of sequence variants is a useful tool in association studies of polygenic traits to identify low-frequency variants and discover novel variants in non-European populations."

While successful, these results suggest that additional associations with height remain to be discovered. Said Du, "we found strong evidence for only one novel height allele, and moderate evidence for another. This emphasizes the need for even larger study populations to discover any remaining associations with rare, but imputable, coding variation related to polygenic traits." Future studies will likely also need to look beyond the exome. Said Du, "our study focused on variants located in or near protein-coding regions. But most (~88%) variants identified from genome-wide association studies have been located in non-coding regions, which suggests these variants may play important regulatory roles and remain important to study." In order to fully characterize the genetic contribution to height, future efforts will need to explore rarer variation within these regions.

Other PHS investigators contributing to this project were Drs. Paul Auer, Shuo Jiao, Christopher Carlson, Cara Carty, Li Hsu, Alex Reiner, Stephanie Rosse, and Charles Kooperberg, as well as Mr. Jeffrey Haessler and Mr. Keith Curtis.

Citation:

[Du M, Auer PL, Jiao S, Haessler J, Altshuler D, Boerwinkle E, Carlson CS, Carty CL, Chen YD, Curtis K, Franceschini N, Hsu L, Jackson R, Lange LA, Lettre G, Monda KL; National Heart, Lung, and Blood Institute \(NHLBI\) GO Exome Sequencing Project, Nickerson DA, Reiner AP, Rich SS, Rosse SA, Rotter JI, Willer CJ, Wilson JG, North K, Kooperberg C, Heard-Costa N, Peters U. 2014. Whole-exome imputation of sequence variants identified two novel alleles associated with adult body height in African Americans. \*Hum Mol Genet\* pii: ddu361. \[Epub ahead of print\]](#)

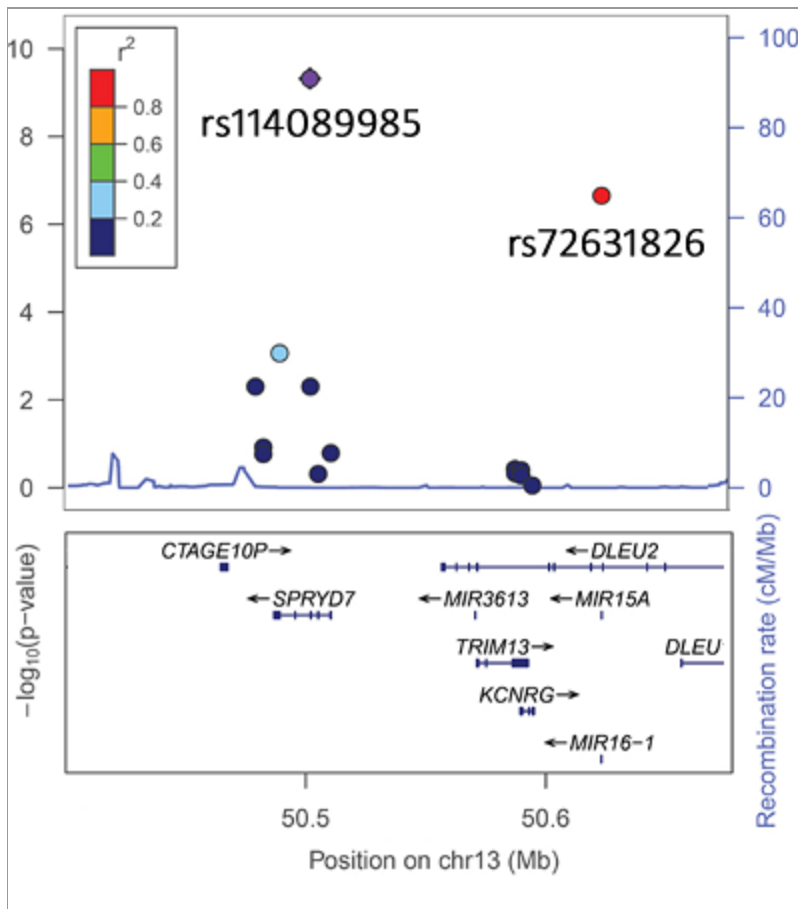


Image provided by Dr. Margaret Du

Regional plot of the associations between adult body height and genetic variants in the 13q14.2 locus, with each dot representing the  $-\log_{10}$  p-value for the association with a particular variant. The color of each dot corresponds with the correlation of that SNP with the index SNP (purple diamond). In conditional analyses, rs72631826 was no longer significant when rs114089985 was included in the model (not shown).